Letters to the Editor

From Mr. Arne Fisher: A Relation Between Two Coefficients in the Gram Expansion of a Function

From Dr. W. A. Shewhart: A Reply

From Mr. Fisher: A Further Note

To the Editor of the Bell System Technical Journal:

In a number of valuable and interesting contributions to this *Journal*, Dr. W. A. Shewhart has made an extended use of the infinite series of Gram. With all the controversy that at present is going on between the pure empiricists, attempting on the one hand to dragoon statistical analysis into a mere *inductio per simplicem enumerationem*, and the a priori theorists on the other hand, who claim that statistical methods so-called are nothing more than simple and evident applications of well-known principles of the probability calculus as formulated by Laplace, it has been a source of satisfaction to me to note that Dr. Shewhart apparently has given the latter methods a place of preference over the methods of the out and out empiricists.

Because of the fact that I happen to be responsible for having called the attention of English-speaking readers to the series of Gram and to have emphasized that Gram's development anteceded the less general developments by Edgeworth and the very special formula by Bowley by more than 20 years, I hope that I may be afforded an opportunity through the medium of your *Journal* to point out in brief form a few decidedly simple features of the Gram series which greatly add to its practical applications in statistical work.

Moreover, it seems that Dr. Shewhart, as well as other students in this country, have received a somewhat different idea about the nature of the Gram series than that which it was my intention to convey in my book on "The Mathematical Theory of Probabilities." This probably is my own fault. For while I have given in the above-mentioned book a description of the various methods for determining the coefficients of the individual terms of the Gram series, I did not mention the various degrees of approximations according to the number of terms as retained in the series itself. The reason for this omission is due primarily to the fact that I expect to treat this aspect in a forthcoming second volume of the book on probability in connection with the presumptive error laws of the a posteriori determined semi-invariants, which laws contain as a special case the evaluation of the standard (or probable) errors of the constants of the frequency curves.

The omission on my part to properly emphasize the close relation between the theory of sampling (i.e., the a posteriori probability theory) and the Gram series is probably also responsible for the fact that Dr. Shewhart in several of his articles has intimated that two terms in the Gram series in certain instances yield a better approximation than three or more terms. This idea has probably arisen from the mistaken notion on the part of Bowley of the generalized probability curve, which is a special example of the general Gram series. The following brief remarks should, therefore, not be taken as a criticism of Dr. Shewhart's work, but rather as a sort of amplification of some of the chapters in my own book on "The Mathematical Theory of Probabilities."

Gram's series, like the Fourier series, offers a perfectly general method for the expansion of arbitrary functions and is, contrary to the opinion of some students, not limited to frequency functions, although it there happens to be especially useful.

The underlying principles of the Gram series may be set forth briefly as follows: Let $F(x)$ be the true (or presumptive) function, which is known from either purely *a priori* considerations, or from observations, and let $G(x)$ be another function (the so-called generating function), which gives a rough approach to $F(x)$. Then according to Gram's method, we have

$$F(x) = c_0G(x) + c_1G'(x) + c_2G''(x) + \ . \ . \ . + c_nG^n(x). \qquad (1)$$

The generating function $G(x)$ may assume a variety of forms. In the case of generalized frequency functions, it is customary to select as the generating function, $G(x)$, a quantity $z = h(x)$ which is normally distributed, and write $F(x)$ as [1]

$$F(x) = c_0\varphi_0(z) + c_1\varphi_1(z) + c_2\varphi_2(z) + \ . \ . \ . + c_n\varphi_n(z), \qquad (2)$$

where $\varphi_0(z) = \dfrac{1}{\sqrt{2\pi}}\, e^{-z^2:2}$ is the generator and $\varphi_1(z),\ \varphi_2(z)\ \ .\ .\ .\ \ \varphi_n(z)$ its derivatives.

When viewed from the theory of elementary errors as originally introduced by Laplace in his monumental work, "Theorie des Probabilities," the Gram series takes on special significance in the way in

[1] If $z = h(x) = (x - M):\sigma$, or a linear function of x, and if the origin of the co-ordinate system is laid at M with σ as its unit, we have the *special* case, or the Charlier A series of the well-known form

$$F(x) = N[\varphi_0(z) + \beta_3\varphi_3(z) + \beta_4\varphi_4(z) + \ . \ . \ . \ ].$$

The various types of the frequency curves of Pearson may of course also be used as generators in the Gram series.

which the possible combinations of the "elementary errors" actually enter into the expansion. It can be shown that there exists a definite relationship between on the one hand the relative order of magnitude of the elementary errors and, on the other, the arrangement of the individual terms of the Gram series.[2]

This relationship was already established by Thiele. It was probably first concisely formulated by Edgeworth, and later on by Charlier and Jörgensen.

The various degrees of approximations can be expressed by the following schemata:

1st approximation $\varphi_0(z)$,
2d approximation $\varphi_0(z) + c_3\varphi_3(z)$,
3d approximation $\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + c_6\varphi_6(z)$,
4th approximation $\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + c_6\varphi_6(z) + c_5\varphi_5(z) + c_7\varphi_7(z) + c_9\varphi_9(z)$.

The first approximation is the usual normal curve. The second is the one which the English statistician, Bowley, erroneously thinks represents a generalized frequency function and for which Dr. Shewhart has shown a marked preference. The third approximation, except for the term involving the sixth derivative, has been used very extensively by Charlier.

Through the publication by C. V. L. Charlier in 1906 of extensive tables to four decimal places of the third and fourth derivatives, the Gram series was made available for practical statistical work in the case of frequency distributions with a moderate degree of skewness and excess (kurtosis). But although Charlier was aware of the fact that the retention of the fourth derivative—which is related to excess (kurtosis)—automatically brings about the inclusion of the sixth derivative, it was not before Jörgensen issued his large numerical tables of the first six derivatives to seven decimal places that we were able to do full justice to the third approximation of the Gram series. Incidentally it might in this connection be mentioned that it is doubtful if the much lauded test for "goodness of fit" as devised by Pearson

[2] Whenever we use the method of moments, the arrangement of the individual terms is *not* arbitrary but must be made according to "order of magnitude" of the various derivatives; and the orders of magnitudes do not correspond to the indices of the derivatives. The generic term "order of magnitude" has in this instance only reference to the formation of the "elementary errors"; if taken in any other sense it is meaningless. The fourth and sixth derivatives are of the same order of magnitude; while the fifth, seventh and ninth all are of the next order following the fourth and sixth. The concept of the different orders of magnitude of the elementary errors is due to Poisson who already in 1832 arrived at the second approximation of the Gram series.

really is able to test the graduating ability of the Gram series as adequately as the more powerful, although far more complicated, "error critique" of Thiele. From Pearson's derivation it appears that his test is not able to take care of elementary errors beyond the first or second order, while it is necessary to consider the formation of elementary errors of the third order in the third approximation of the Gram series. In some work I have been doing in the way of construction of compound mortality curves, I have at least found that the Pearson test is inadequate, if actually not misleading, because it apparently fails to measure the effect of the elementary errors of higher order which enter into the formation of such compound mortality curves.

There exists, however, a very simple relationship between the coefficients c_3 and c_6 in the third approximation. We have, namely, with a fair approach to exactitude, the simple relation: $c_6 = \frac{1}{2}c_3^2$. It is therefore not necessary to calculate the semi-invariants or moments of higher orders than those of the fourth order, since we shall have

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + \tfrac{1}{2}c_3^2\varphi_6(z)$$

as a third approximation.

As an illustration of the above formula, we may select the expansion of the point binomial $(0.1 + 0.9)^{100}$. We have here, according to the formulas on pages 263–264 of my "Mathematical Theory of Probabilities":

$$s = 100, \qquad p = 0.1, \qquad q = 0.9$$

and

$$\lambda_1 = M = sp = 10, \qquad \sigma = \sqrt{spq} = 3, \qquad c_3 = -0.0444, \qquad c_4 = 0.0021$$

and

$$c_6 = \tfrac{1}{2}c_3^2 = 0.0010,$$

or

$$(0.1 + 0.9)^{100} = \tfrac{1}{3}[\varphi_0(z) - 0.0445\varphi_3(z) + 0.0021\varphi_4(z) + 0.0010\varphi_6(z)],$$

where

$$\varphi_0(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2:2}$$

and

$$z = (x - 10) : 3.$$

A comparison between the above approximation and the true expansion of the point binomial $(0.1 + 0.9)^{100}$ to 4 decimals is given in the following table.

x = No. of Successes	Gram Series	True Value	No. of Successes	Gram Series	True Value
0	.0000	.0000	13	.0744	.0743
1	.0003	.0003	14	.0515	.0513
2	.0016	.0016	15	.0327	.0327
3	.0060	.0059	16	.0192	.0193
4	.0160	.0159	17	.0105	.0106
5	.0338	.0339	18	.0054	.0054
6	.0594	.0596	19	.0026	.0026
7	.0888	.0889	20	.0012	.0012
8	.1149	.1148	21	.0005	.0005
9	.1305	.1304	22	.0002	.0002
10	.1318	.1319	23	.0001	.0001
11	.1198	.1199	24	.0000	.0000
12	.0988	.0988			

The approximation is in this case well nigh perfect and comes much closer to the true values of the point binomial than any of the six approximations as given in Dr. Shewhart's article in the January 1924 number of this *Journal*. It also shows that with exactly the same amount of computation as that involved in the so-called Charlier A series, we can reach greatly improved results through the inclusion of the sixth derivative in the series. This arises from the important fact that once we have computed the coefficients c_3 and c_4, it is not necessary to calculate c_6 since $c_6 = \frac{1}{2}c_3^2$ approximately. Moreover, since extensive tables, notably those of Jörgensen, now are available for the normal function and its first six derivatives, there seems no good reason why we should not use the more exact approximation than the inexact formula by Bowley.

In conclusion, it might be well to emphasize the fact that while it is important to consider the relative order of magnitudes of the separate terms in the Gram series when we use the methods of semi-invariants or of moments, such restrictions are not necessary if we use the method of least squares in conjunction with properly determined weights.

ARNE FISHER.

December 10, 1926.

To the Editor of the Bell System Technical Journal:

I have read Mr. Fisher's communication with considerable interest. We who do not read the Scandinavian language owe much to him for his very able amplification and interpretation of many important contributions of the Scandinavian school of mathematical statisticians and this debt has been increased by the above communication insofar as it brings to light a very interesting relationship (the discovery of

which is attributed to Thiele), namely, that in the notation of the communication the constant c_6 is approximately equal to $\frac{c_3^2}{2}$.

Mr. Fisher definitely states that no criticism of my work is intended, but incidental to bringing out the above relationship he makes certain statements upon which I should like to comment briefly.

He states that the omission on his part to properly emphasize a close relation between the theory of sampling and the Gram series is probably responsible for the fact that I have intimated that two terms of the Gram series in certain instances yield a better approximation than three or more terms. To my knowledge this is not the case.

The special form of the Gram series used in my published articles in this *Journal* is that represented by his Equation 2.[1] The validity of this expansion rests upon the Lebedeff theorem.[2] So far as I am aware I have not intimated that two terms of the series yield a better approximation than three or more terms in the sense that

$$| F(z) - [c_0\varphi_0(z) + c_3\varphi_3(z)]|$$

should be less than

$$| F(z) - [c_0\varphi_0(z) + c_3\varphi_3(z) + \dots + c_n\varphi_n(z)]|$$

irrespective of n, although it is in this sense that Mr. Fisher discusses his example of the graduation of $(.9 + .1)^{100}$. To have done so would have been an obvious blunder because, assuming the Lebedeff theorem to be true, the absolute value of the difference ϵ between the function $F(z)$ and the sum of the first n terms of the series can be made as small as we please by taking n sufficiently large.[3]

I did say, however, in my article in the October issue of this *Journal:* "Carrying out steps 1 and 2, we conclude that the best theoretical equation representing the data in Fig. 1 is either the Gram-Charlier series (2 terms) or the Pearson curve of Type IV for both of which the estimates of the parameters may be expressed in terms of the first four moments μ_1, μ_2, μ_3 and μ_4 of Fig. 3." Of course the first two terms of the Gram-Charlier series requires only μ_1, μ_2 and μ_3. "Best" as used here obviously is in the sense of probability of fit which is entirely different from saying that the first two terms is the best approximation in the sense discussed by Mr. Fisher at least as illustrated by his

[1] It is of course understood that, in practice, transformations are made so that c_1 and c_2 are both equal to zero. In what follows, therefore, the second term of the series will be $c_3\varphi_3(z)$.

[2] Fisher, Arne, "Mathematical Theory of Probabilities," 2d edition, 1922, p. 203.

[3] It can be seen from my published work, however, that the sum of two terms is sometimes better than the sum of three.

example. In this case I found that the probability of fit for two terms was greater than that for three. Now, I find that it is as good as for Mr. Fisher's third approximation. It may be of interest also to know that statistical distributions sometimes arise where the first three terms give as good a fit as Mr. Fisher's third approximation involving 4 terms. This is particularly true when the universe from which the sample is drawn is nearly symmetrical. My action in this connection can be justified both upon theoretical and practical grounds but we need not do more than mention this point to make sure that the reader will not confuse my statement quoted above with what Mr. Fisher is talking about in his communication.

Having thus dismissed the questions which may arise in connection with published work in this *Journal*, I should like to add a word or two of caution to the reader of Mr. Fisher's letter where it reads: "Moreover, since extensive tables, notably those of Jörgensen, now are available for the normal function and its first six derivatives, there seems no good reason why we should not use the more exact approximation than the inexact formula by Bowley."

We have made far more use of the Gram series in connection with our inspection work than indicated in the published papers. In this work we have found that it is theoretically not necessary in certain instances and in many more instances it is not practical to follow Mr. Fisher's suggestion. I shall limit my remarks to the application of the series which we have made in expanding a known function in terms of an infinite series in which the generating function is the normal law. In this connection the outstanding practical question is: Given the known function $F(x)$, what number n of terms of the infinite series must we take in order that the absolute magnitude of the difference between the function $F(x)$ and the sum of the n terms will be less than a given preassigned quantity ϵ? I am sorry that Mr. Fisher does not answer this question. Instead he proposes a grouping of terms upon the basis suggested in a footnote to his article. Now, it may easily be shown in the particular case cited by Mr. Fisher, i.e., the graduation of the point binomial $(.9+.1)^{100}$, that the sequence of signs depends upon the value of z, that for certain values of z his second approximation is just as good as his third, and that in many instances the difference between the second approximation and the third is not sufficiently great to be of any practical importance. Whether we should use the second, third, or higher approximation in a given case is one for special consideration.

In closing let me say that I have not made the above remarks with any intention of discrediting the applications of this series but rather

to indicate to the casual reader that there are certain technical questions involved in its application which must be given due consideration even beyond the stage outlined in Mr. Fisher's communication. I have found that this series often has many advantages over competing methods of analyzing data although not all of these advantages are referred to in the literature of the subject.

W. A. SHEWHART.

December 28, 1926.

To the Editor of the Bell System Technical Journal:

The question raised by Dr. Shewhart as to the measure of the absolute magnitude of the difference between a known function, $F(x)$, and the first n terms of the series has been treated by Gram in his original article on "*Rækkeudviklinger bestemte ved Hjælp af de mindste Kvadraters Metode.*" (On Development of Series by means of the Method of Least Squares.) In this article Gram also discusses at length the decidedly practical question of arriving at an estimate of the remainders (or residuary terms), which invariably occur in practice where we, of course, are forced to deal with a finite number of terms.

It would, however, be beyond the limits of the present communication to enter into this aspect of the question, which necessarily is somewhat complicated. In passing it, I wish merely to state that Gram's original method of determining the coefficients in the series on the basis of the principle of least squares is decidedly easier to apply than the relatively cumbersome method of moments in arriving at a reliable measure of the remainder of the series after, say, the n^{th} term.

Dr. Shewhart's further contention that two terms of the Gram series sometimes give as good a fit as three or even four terms, and that three terms in the case of nearly symmetrical distributions serves as well as four terms, seems to me to be almost self-evident from a simple consideration of the way in which the coefficients c actually enter into the series.

All the terms containing uneven indices tend to produce skewness, and all the terms with even indices produce excess (kurtosis). If the coefficient c_3 is not too large, and if c_4 is small as compared with c_3, it is evident that

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z)$$

will give about as good an approximation as

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + \tfrac{1}{2}c_3{}^2\varphi_6(z).$$

On the other hand, in nearly symmetrical distributions with a pro-

nounced excess (kurtosis), where c_4 is large as compared with c_3, it seems also reasonable that

$$F(x) = c_0\varphi_0(z) + c_4\varphi_4(z)$$

might in certain instances give as good a fit as

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + \tfrac{1}{2}c_3{}^2\varphi_6(z).$$

These aspects of the series have been discussed by Thiele.

ARNE FISHER.

January 10, 1927.